

# Evaluating gene regulatory network reconstruction methods when post-transcriptional modifications are present

Olivia Angelin-Bonnet

Massey University

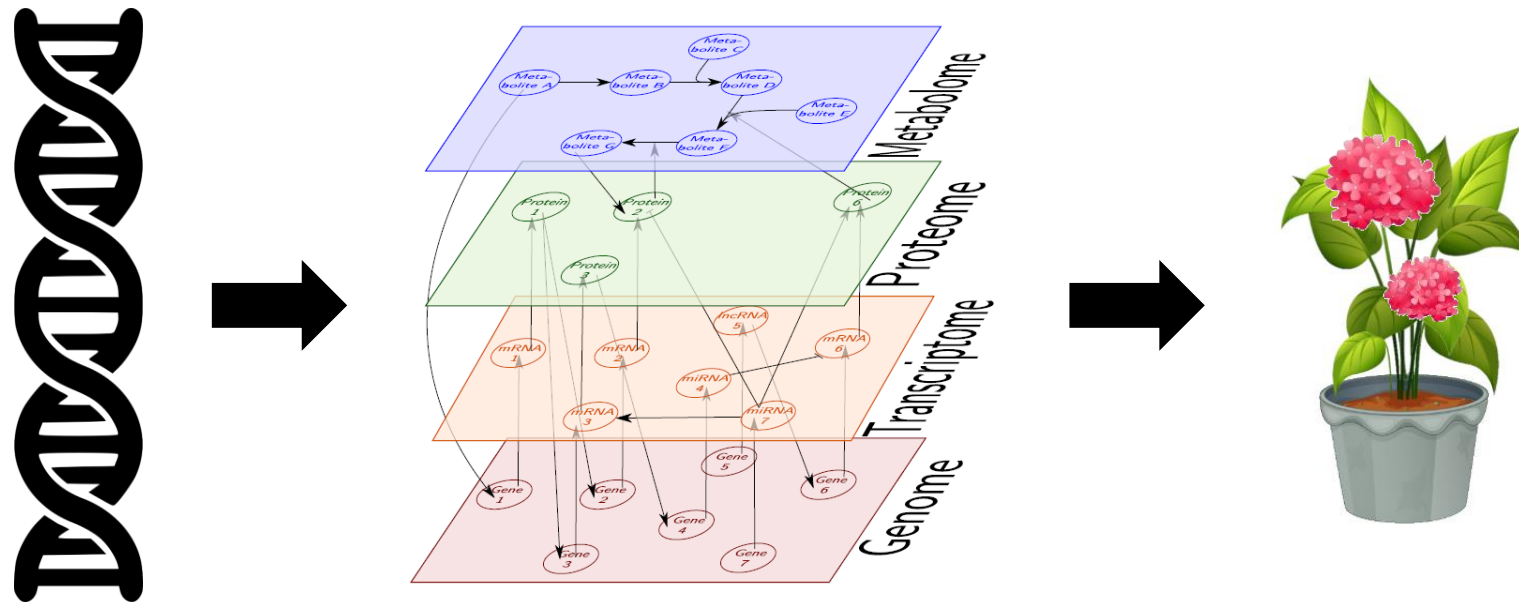
20 September 2021



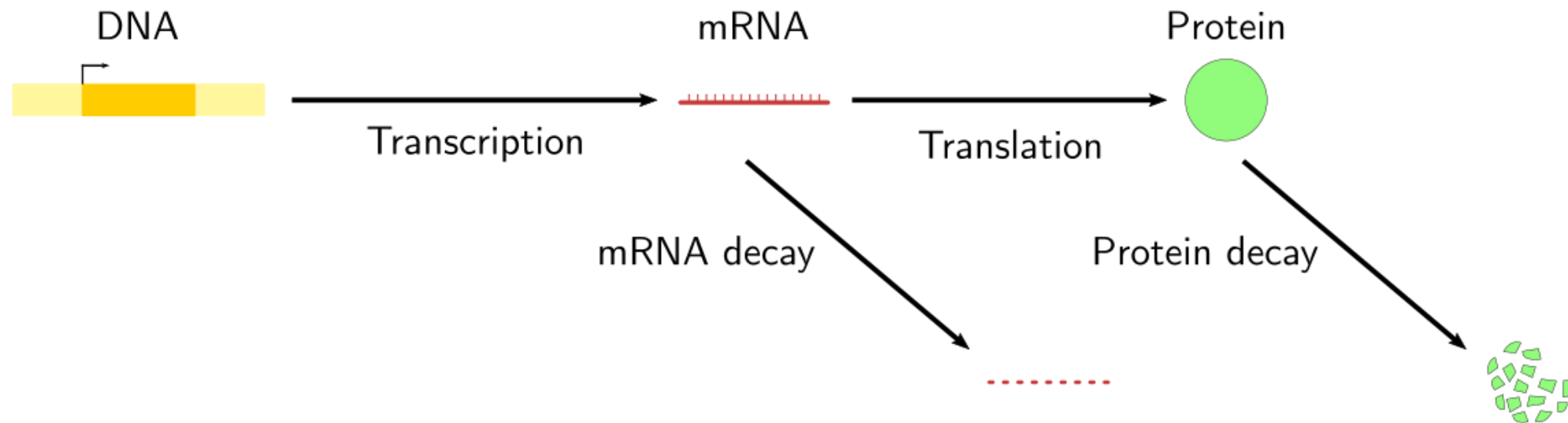
MASSEY UNIVERSITY  
TE KUNENGA KI PŪREHUROA  
UNIVERSITY OF NEW ZEALAND

Dr Matthieu Vignes  
A/Prof Patrick Biggs

# Introduction – from genotype to phenotype



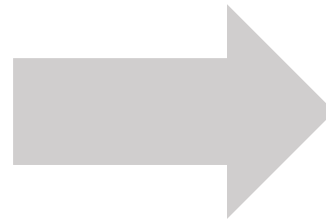
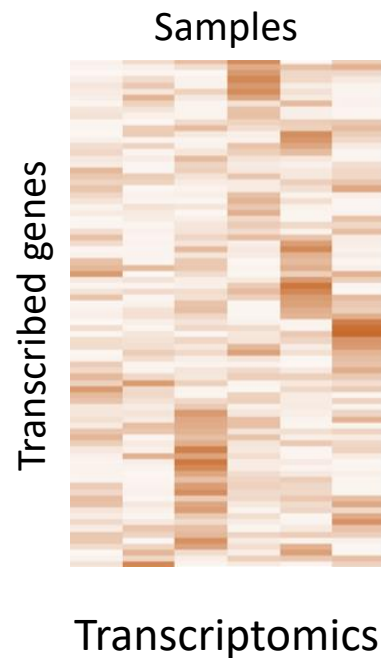
# Introduction – post-transcriptional regulation



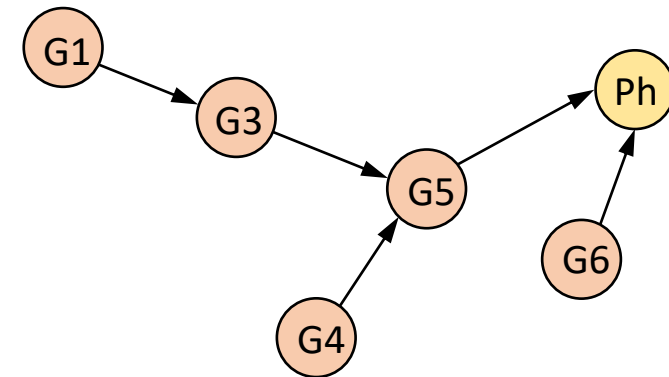
Adapted from Angelin-Bonnet *et al.*, Humana Press (2019)

# Introduction – GRN reconstruction

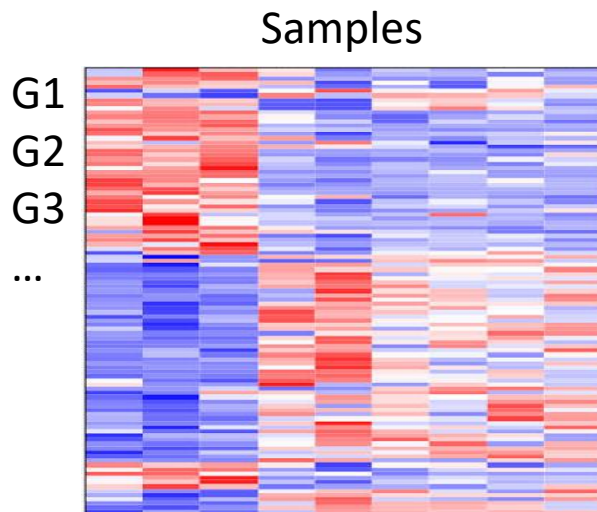
Gene expression data



Causal  
gene regulatory network  
(GRN)



# Introduction – causal inference methods



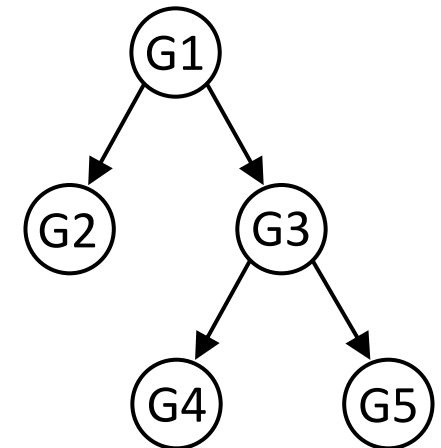
Observational data

$$G2 \perp\!\!\!\perp G3 \mid G1$$

$$G1 \perp\!\!\!\perp G5 \mid G3$$

...

Set of conditional independence between features



Causal graph

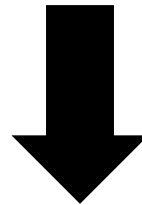


# Introduction – the research question

Can causal inference methods detect post-transcriptional regulation from transcriptomics data?

# Introduction – the research question

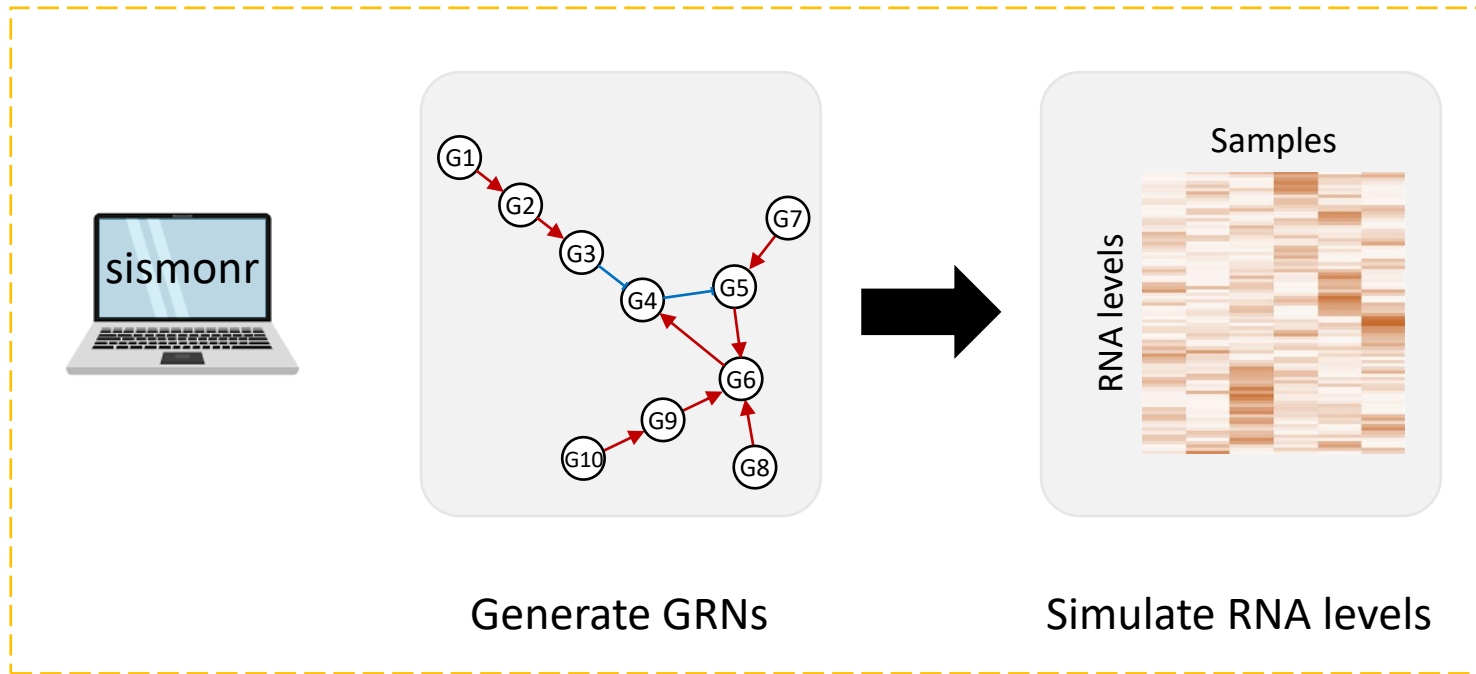
Can causal inference methods detect post-transcriptional regulation from transcriptomics data?



Evaluation of causal inference methods on simulated datasets



# Step 1: simulations





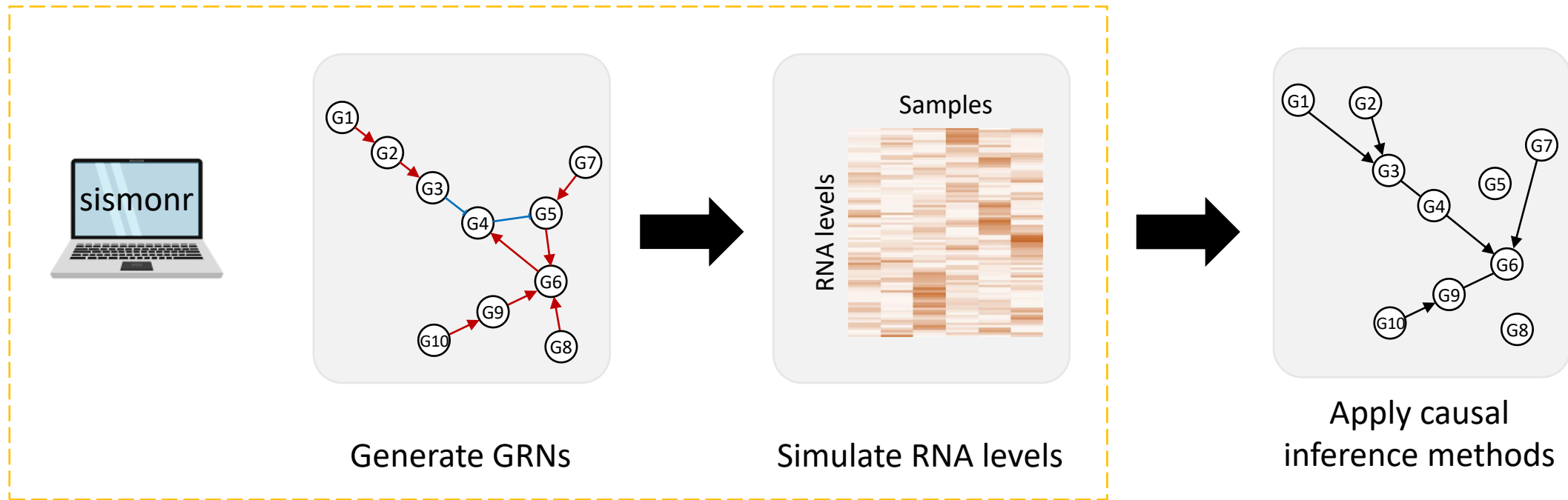
# Simulation configurations

Configuration	Number of TC* regulators	Type of post-TC* regulation	Number of post-TC* regulators
Configuration 1	10	-	0
Configuration 2	7	Translation	3
Configuration 3	7	RNA decay	3
Configuration 4	7	Protein decay	3
Configuration 5	7	Protein PTM <sup>†</sup>	3
Configuration 6	5	Translation	5
Configuration 7	5	RNA decay	5
Configuration 8	5	Protein decay	5
Configuration 9	5	Protein PTM <sup>†</sup>	5
Configuration 10	3	Translation	7
Configuration 11	3	RNA decay	7
Configuration 12	3	Protein decay	7
Configuration 13	3	Protein PTM <sup>†</sup>	7

\* TC = transcription

† PTM = post-transcriptional modification

# Step 2: causal inference

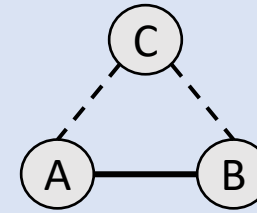


# The causal inference methods

## Constraint-based methods

Conditional independence tests

e.g.: PC, FCI



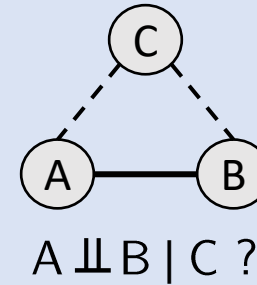
$A \perp\!\!\!\perp B \mid C ?$

# The causal inference methods

## Constraint-based methods

Conditional independence tests

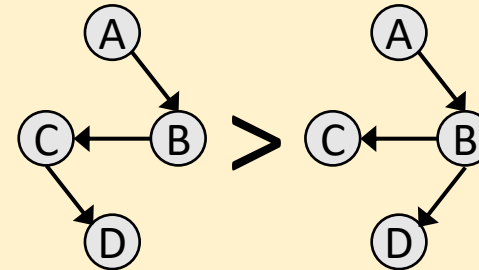
e.g.: PC, FCI



## Score-based methods

Candidate networks scoring

e.g.: GES

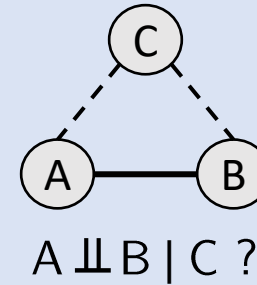


# The causal inference methods

## Constraint-based methods

Conditional independence tests

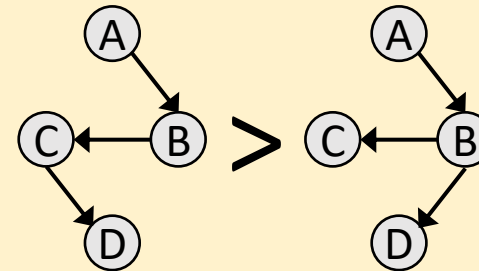
e.g.: PC, FCI



## Score-based methods

Candidate networks scoring

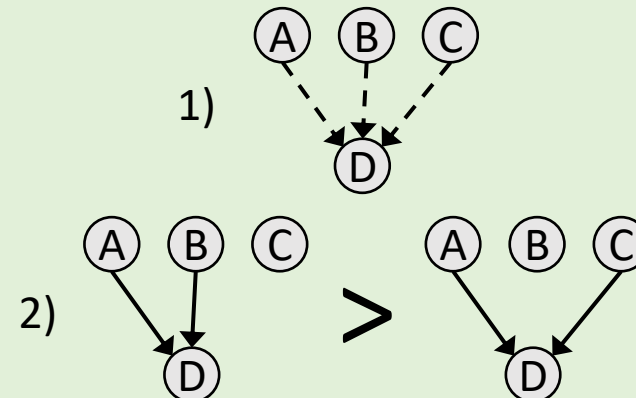
e.g.: GES



## Hybrid methods

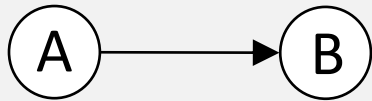
Use both approaches

e.g.: MMHC, SC



# Types of causal graphs constructed

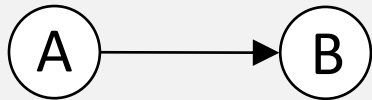
## Directed Acyclic Graph (DAG)



Causal relationship **from A to B**

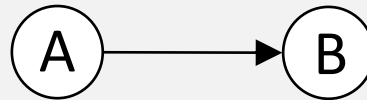
# Types of causal graphs constructed

## Directed Acyclic Graph (DAG)

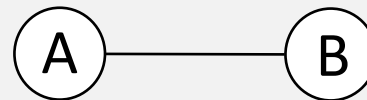


Causal relationship **from A to B**

## Completed Partially Directed Acyclic Graph (CPDAG)



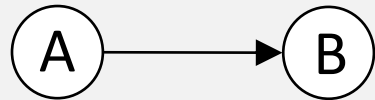
Causal relationship **from A to B**



Causal relationship between A and B, **directionality undetermined**

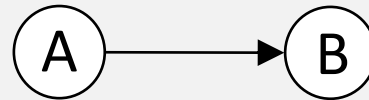
# Types of causal graphs constructed

## Directed Acyclic Graph (DAG)

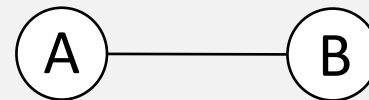


Causal relationship **from A to B**

## Completed Partially Directed Acyclic Graph (CPDAG)

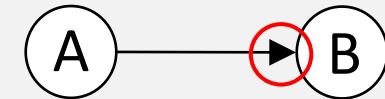


Causal relationship **from A to B**

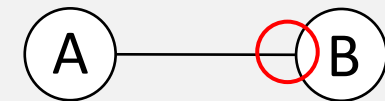


Causal relationship between A and B, **directionality undetermined**

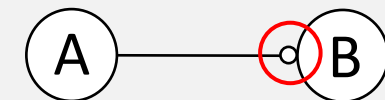
## Partial Ancestral Graph (PAG)



**No ancestral relationship from B to A**



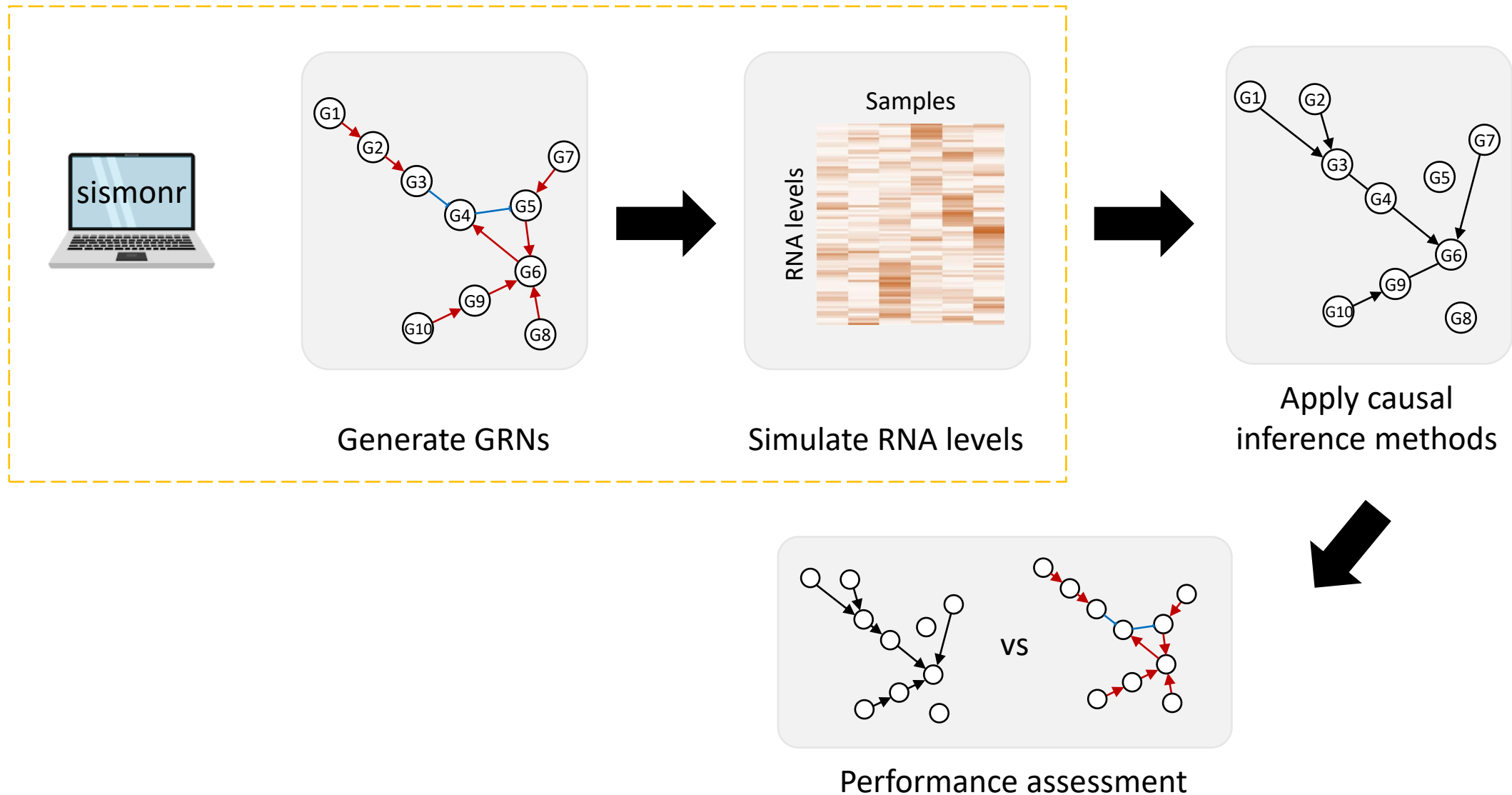
**Ancestral relationship from B to A**



**Directionality undetermined**

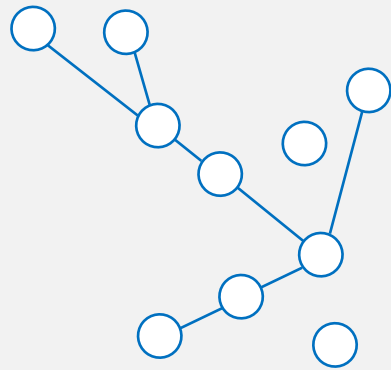


# Step 3: performance assessment



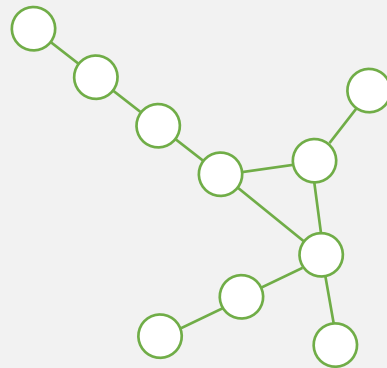
# Evaluation of causal graphs

## Undirected networks



Reconstructed network

VS



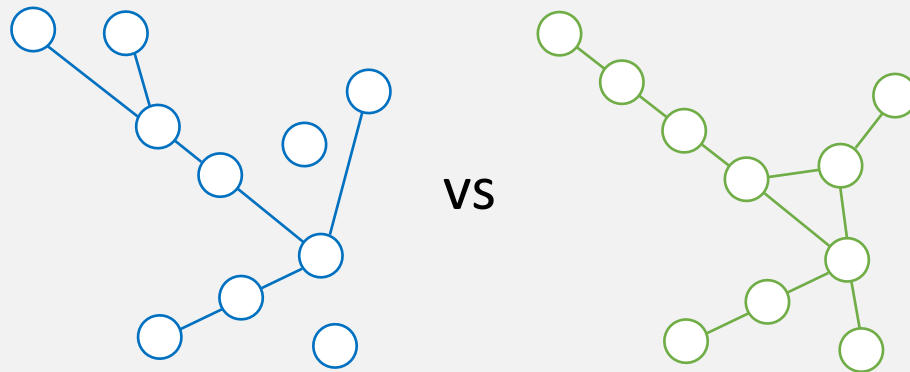
True network

True Positive	False Positive
True Negative	False Negative

ROC / PR curves

# Evaluation of causal graphs

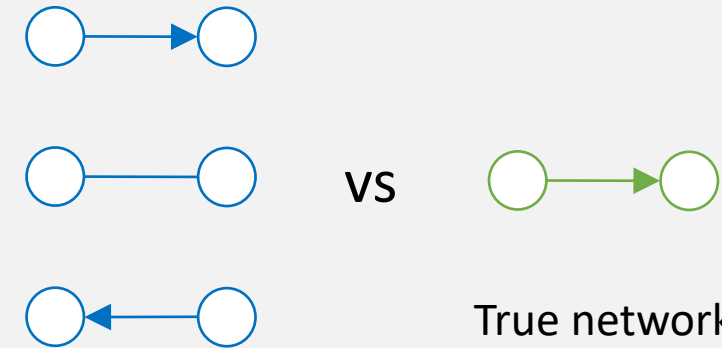
## Undirected networks



Reconstructed network

True network

## Partially directed networks



Reconstructed network

True network

True Positive	False Positive
True Negative	False Negative

ROC / PR curves

- Scores
- Causal queries (Heinze-Deml *et al.*, 2018)



# Causal queries

For a given causal graph, and pair of nodes A and B:

**Is A a causal parent of B?**



# Causal queries

For a given causal graph, and pair of nodes A and B:

**Is A a causal parent of B?**

For a DAG:

Yes if 

For a CPDAG:

Yes if 

For a PAG:

Cannot answer

# Causal queries

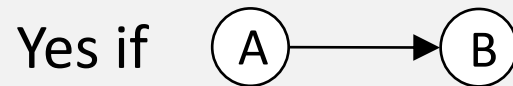
For a given causal graph, and pair of nodes A and B:

**Is A a potential causal parent of B?**

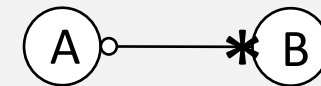
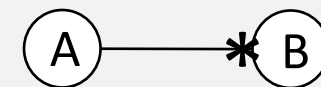
For a DAG:



For a CPDAG:



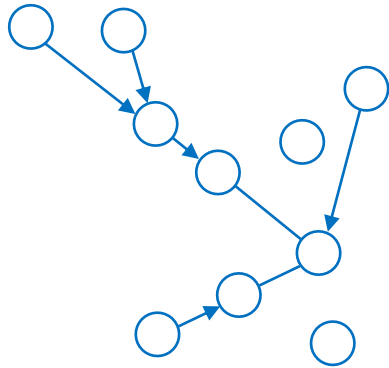
For a PAG:



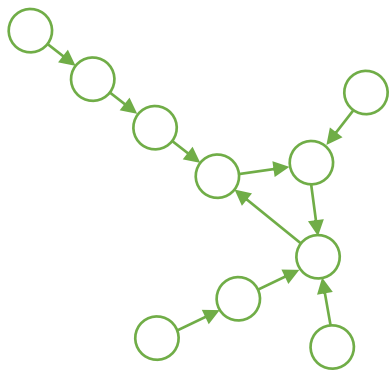
# Causal queries

Query	DAG	CPDAG / undirected graph	PAG
$A$ parent of $B$	$A \rightarrow B$	$A \rightarrow B$	$\emptyset$
$A$ potential parent of $B$	$A \rightarrow B$	$A \rightarrow B$ $A \text{ --- } B$	$A \text{ ---}^* B$ $A \circ \text{---}^* B$
$A$ not parent of $B$	Complement of potential parent query		
$A$ ancestor of $B$	path from $A$ to $B$ with edges $A \rightarrow B$	path from $A$ to $B$ with edges $A \rightarrow B$	path from $A$ to $B$ with edges $A \text{ ---}^* B$
$A$ potential ancestor of $B$	path from $A$ to $B$ with edges $A \rightarrow B$	path from $A$ to $B$ with edges $A \rightarrow B$ and $A \text{ --- } B$	path from $A$ to $B$ with edges $A \text{ ---}^* B$ and $A \circ \text{---}^* B$
$A$ not ancestor of $B$	Complement of potential ancestor query		

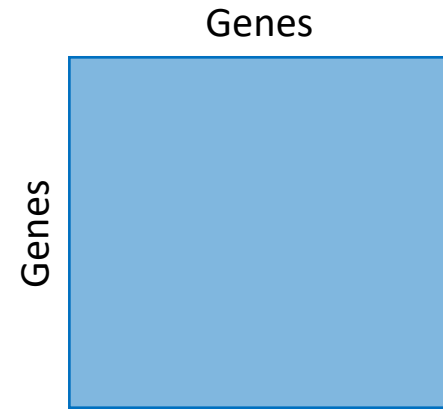
# Scoring causal graphs based on causal queries



Answer causal query on reconstructed network



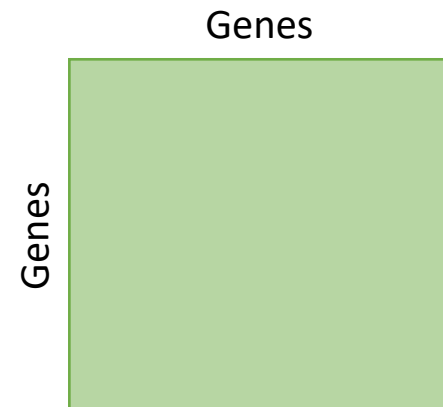
Answer causal query on true network



Binary answer matrix



True Positive	False Positive
True Negative	False Negative



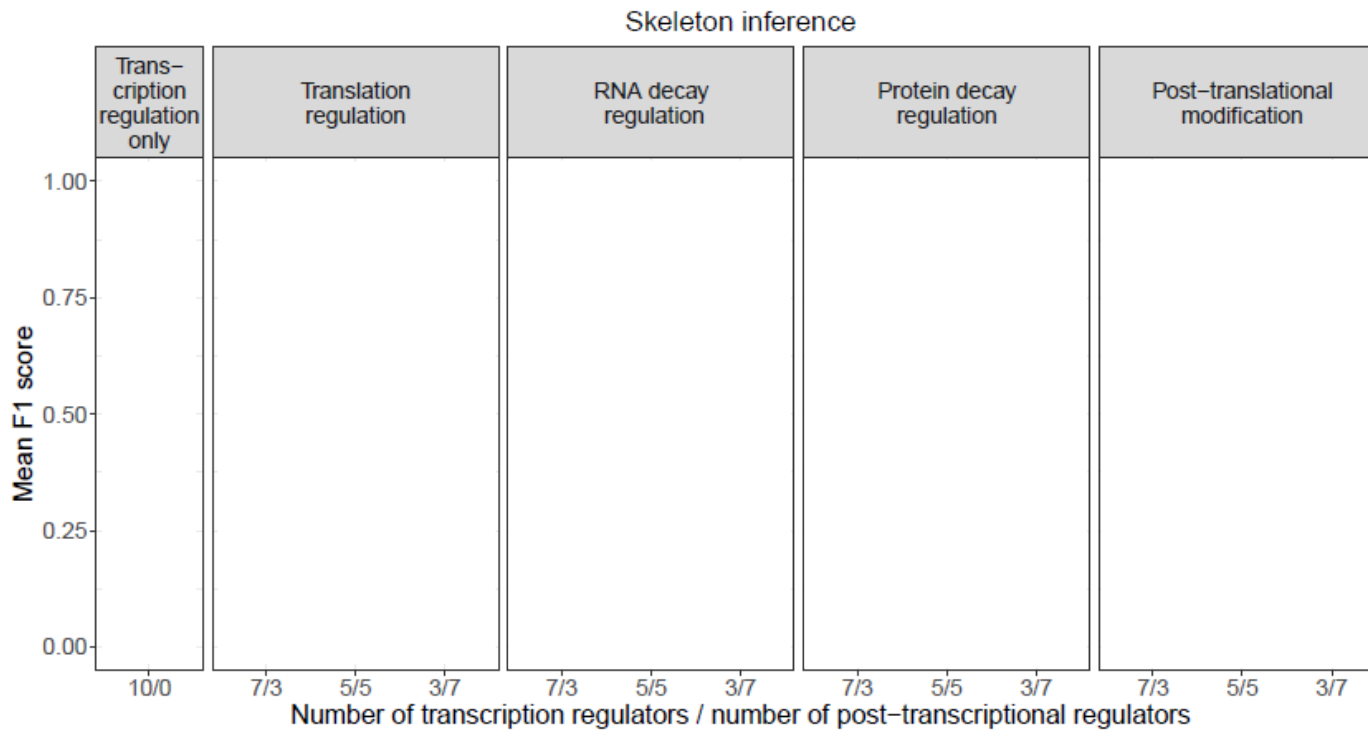
Binary answer matrix



- Precision
- Recall
- $F_1$ -score

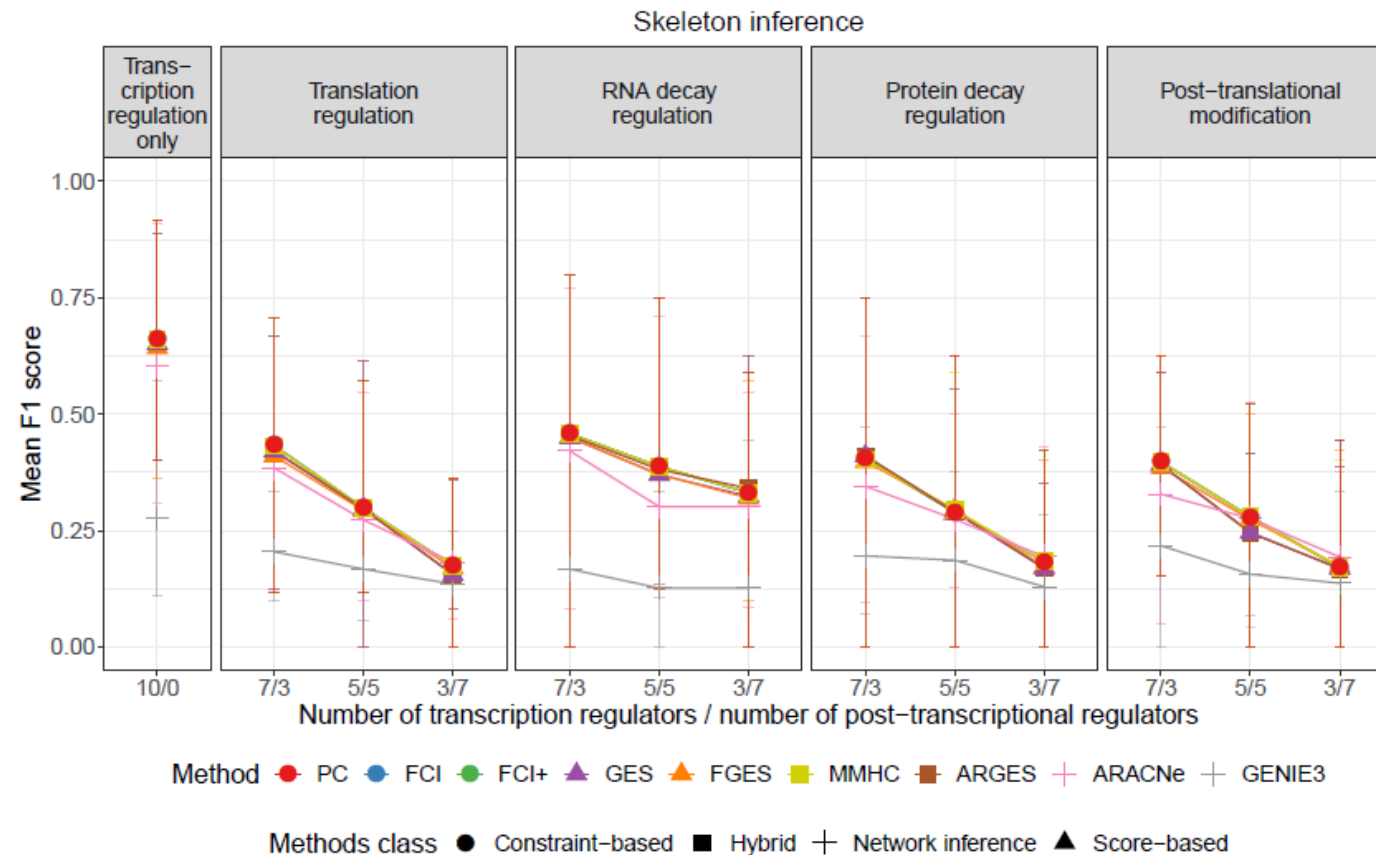


# Performance assessment: (some) results

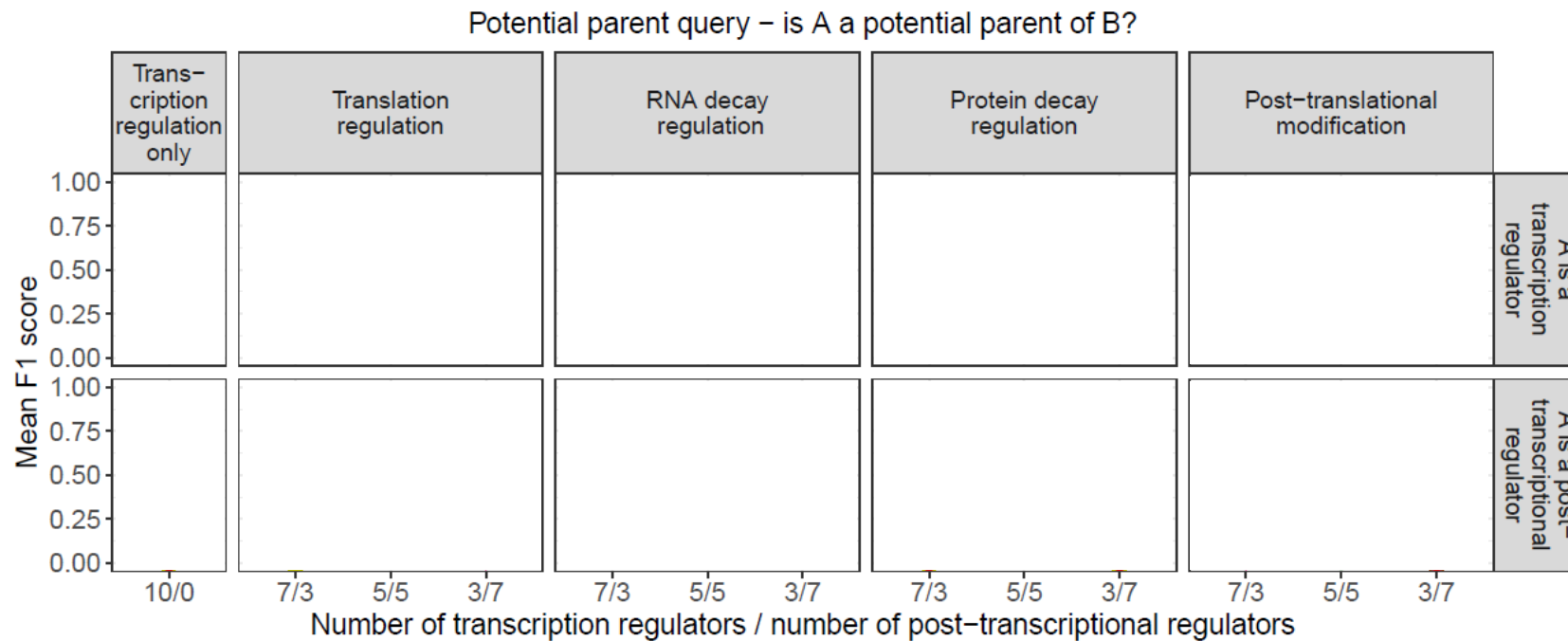


# Performance assessment: (some) results

Causal inference methods are ok to detect the presence of causal relationships...  
as long a post-transcriptional regulation is not present



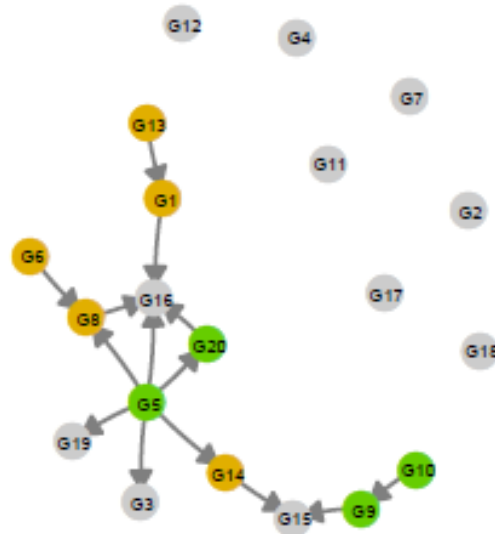
# Performance assessment: (some) results





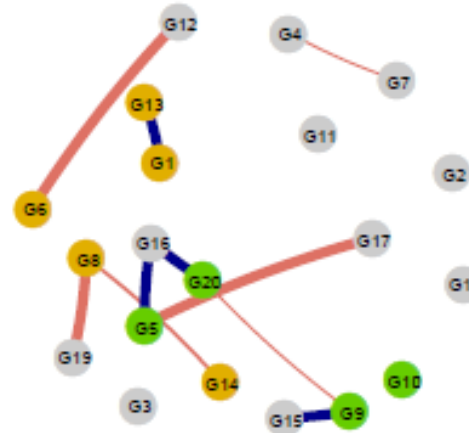
# Performance assessment: (some) results

True GRN

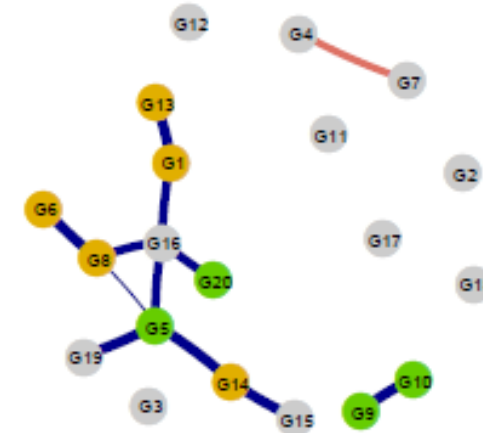


Reconstructed causal graph based on

RNA measurements



Protein measurements



Nb. of methods that inferred the edge:

— 1 — 2 — 3 — 4 — 5 — 6

Is the edge in the true skeleton:

— No — Yes

Regulator of:

● Transcription ● Translation ● - (target gene)



# Next steps

## Other questions to answer:

- How resilient are the methods to violation of their assumptions? (e.g. feedback loops, cycles)
- Are some motifs easier to detect?
- How do these methods scale up (i.e. for larger networks)?

# Conclusion

- Causal inference methods to go beyond gene association networks
- Need to account for post-transcriptional regulation... using complementary datasets?
- Causal queries very useful to compare (partially) directed networks...
- ... and to make sense of directed biological networks

# Thank you for your attention!

Any questions?



**MASSEY UNIVERSITY**  
**TE KUNENGA KI PŪREHUROA**  
**UNIVERSITY OF NEW ZEALAND**